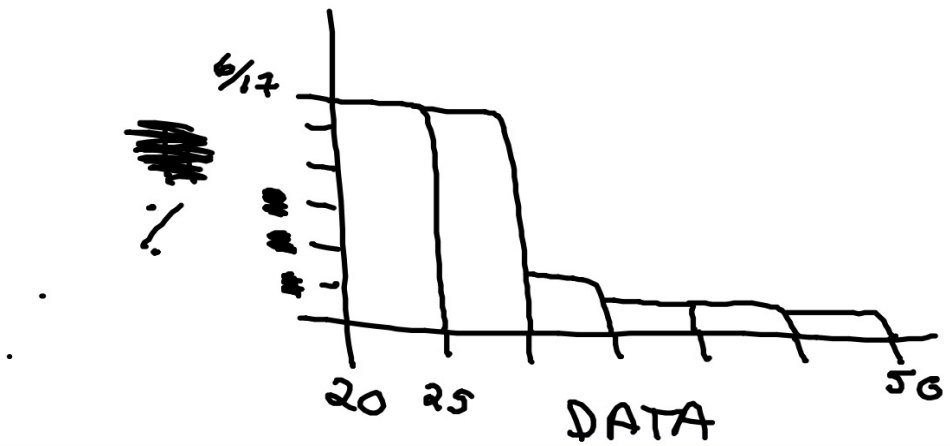
Review: histograms

Create a histogram of the following data, then describe the distribution.

20	20	21	23	24	22	25
25	25	25	26	28	30	34
38	40	46				



MEDIAN: Center

- put numbers in order
- find the middle number
- if there are two middle numbers, average the two

Example:

3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20

3 3 4 5 6 6 ~~7~~ 8 10 11 12 15 18 20

7.5 = Med.

QUARTILES:

- the medians of the lower and upper halves of the data
- first and third quartiles
- Q1 and Q3

Q2 = Median

25%

Same example:

7.5

3, 3, 4, 5, 6, 6, 7, 8, 10, 11, 12, 15, 18, 20

Q1 = 5

Q3 = 12

IQR: Inter-quartile range

- the difference between Q3 and Q1 (a, b)

IQR = Q3 - Q1

- the middle 50% of the data

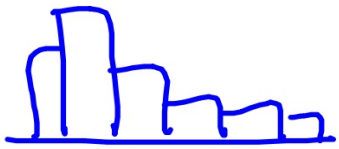
Same example:
Q3 = 5
Q1 = 12

IQR = 12 - 5 = 7 (3, 20)

THE 5# SUMMARY:

Ex:

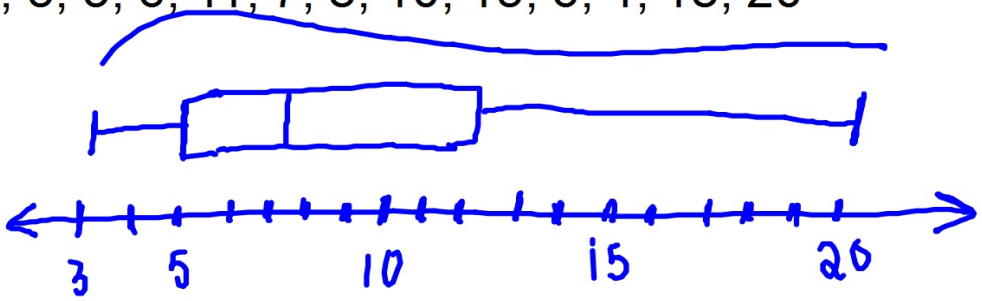
Min = 3
Q1 = 5
Median = 7.5
Q3 = 12
Max = 20



Visually: BOXPLOT Box & Whisker plot

Use the data from before:

3, 12, 5, 3, 6, 11, 7, 8, 10, 15, 6, 4, 18, 20



5 # Summary and Boxplots on the calculator:

Find the 5# Summary & create a boxplot of the following data:

12	18	25	20	21	17	30	29	32	35
10	40	38	34	33	31	30	25	24	23
19	22	13	16	26	29	27	35	62	57

Finding outliers- the 1.5 x IQR test

In order to formally to determine if there are outliers:

- find IQR $= Q_3 - Q_1$
- multiply by 1.5
- Take this number and:
    - subtract from Q1
    - add to Q3
- This is considered the range of acceptable data
- ANYTHING OUTSIDE THIS RANGE IS CONSIDERED AN OUTLIER

Example: Using the data from before, test for outliers

12	18	25	20	21	17	30	29	32	35
10	40	38	34	33	31	30	25	24	23
19	22	13	16	26	29	27	35	62	57

min = 10
Q1 = 20
M = 26.5
Q3 = 33
Max = 62

IQR = 13
13 × 1.5 = 19.5

20 − 19.5 = 0.5
33 + 19.5 = 52.5

2 outliers: 57, 62

Modified Boxplot:

- If you decide that there are outliers, you can modify your boxplot

- Put a dot where the outliers are, and then make the whisker go to the next available point

- Try this with the example data from before

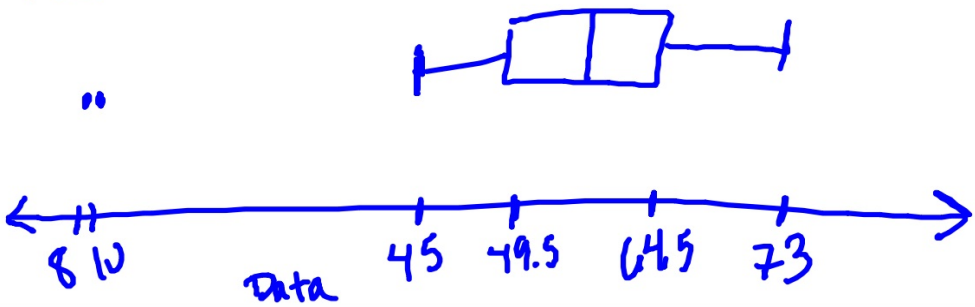
Example: Create a boxplot (testing for outliers) of the following data:

10	45	60	62	55	50	49	51	52	55	56
69	67	63	64	66	65	70	72	73	48	47
46	47	53	54	57	58	59	60	70	8	60

min = 8
Q1 = 49.5
M = 57
Q3 = 64.5
max = 73

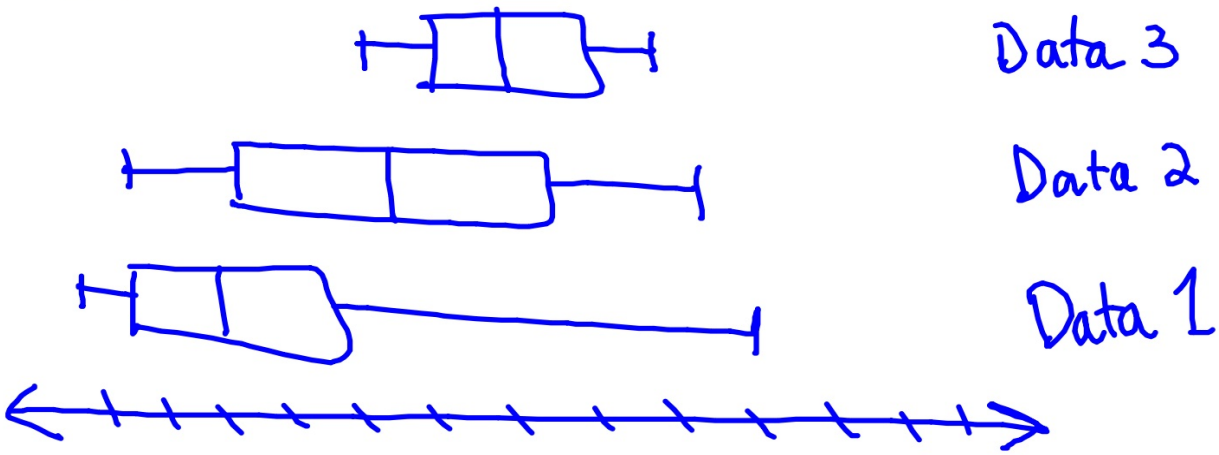
IQR = 15 x 1.5 = 22.5

$Q_1 - 22.5 = 27$
$Q_3 + 22.5 = 87$



Parallel boxplots

- Compare 2 or more sets of data



Start HW!!